

Egy orosz nyelvű korpusz (NarRu) narratívaelemzése saját fejlesztésű szentiment- és emóciósztárakkal

Nyíri Zsófi¹, Szabó Martina Katalin^{1,2,3}, Ilyés Virág¹

¹PrecognoX Informatikai Kft.

² Szegedi Tudományegyetem, Szlav Intézet, Orosz Filológiai Tanszék

³ MTA TK „Lendület” RECENS Kutatócsoport

nyirizsofi93@gmail.com

{mszabo, vilyes}@precognoX.com

Kivonat: A tanulmány egy orosz nyelvű szentiment- és emóciólexikon, valamint egy szépirodalmi korpusz létrehozásának tapasztalatairól és felhasználásának első eredményeiről számol be. A korpusz és az elemzési eszközök létrehozásának elsődleges célja az volt, hogy a szövegek emóció- és szentimentelemzésével feltárjuk azok narratívastuktúráit. A tanulmányban bemutatjuk a korpusz felépítését és létrehozásának a legfontosabb szempontjait, valamint a szótárak készítésének eszközeit, módszereit, alapelveit és adatait. Ezt követően ismertetjük a korpuszszövegek emóció- és szentimentelemzési eredményeit. Áttekintést adunk arról a megoldásról, amellyel a kapott elemzési eredmények alapján a szövegek narratívastuktúráit feltártuk, és a leggyakoribb narratívastípusokat definiáltuk. Végezetül bemutatjuk azt az interaktív dashboardot, amelynek segítségével a korpuszt a létrehozott eszközökkel a szentimentek, az emóciók, valamint a narratívastuktúrák szempontjából vizsgálhatóvá tettük.

1 Bevezetés

A dolgozatban egy olyan kutatásról számolunk be, amelynek célja különböző orosz nyelvű szépirodalmi szövegek narratívastuktúrájának a feltárása szentiment- és emóciósztárak segítségével. A kutatáshoz a vizsgálati anyagot és az elemzési eszközöket magunk hoztuk létre [1].

A munka egyik fontos célja az volt, hogy olyan szótárakat és korpuszt fejlesszünk, amelyek segítségével szolgálhatnak az orosz nyelvre irányuló emóció-, szentiment-, valamint narratívaelemzési feladatainkhoz. A kutatás alapjául vett tanulmányban Reagan és szerzőtársai [2] szentiment- és emócióelemzéssel vizsgálták angol nyelvű szövegek narratíváit, ami alapján azt feltételeztük, hogy a korpusz szövegeinek szentiment- és emócióelemzési eredményeivel rá tudunk mutatni az egyes szövegek narratívastuktúráira, illetve sikerül erre a szövegsajátságra vonatkozóan típusokat meghatározni.

Annak céljából, hogy különböző korszakokban keletkezett irodalmi művek stuktúráit összevethessük egymással, a vizsgálati korpuszban három alkorpuszt különítettünk el. A kutatás első fázisában elvégzett szótárak elemzéssel megkaptuk a

korpuszban reprezentált korszakokra jellemző szisztematikus mintázatokat az emóció- és a szentimenttartalmak vonatkozásában. Ezt követően, az eredmények alapján feltártuk a szövegek narratívastruktúráit, a különböző szerkezetítípusok korpuszbeli kvantitatív adatait, és definiáltuk a fő narratívátípusokat.

A dolgozatban ismertetjük a vizsgálati korpusz létrehozásának a menetét, módszereit, alapelveit, a vizsgálati eszközök adatait, valamint szólunk a fejlesztés során felmerülő elméleti és gyakorlati dilemmákról is. Ezt követően ismertetjük az elvégzett elemzési megoldásokat, majd a feldolgozással kapott eredményeket mind a szentiment- és emóciótartalmakra, mind a narratívastruktúrákra vonatkozóan, és bemutatjuk az azokat prezentáló interaktív dashboardot. Végül szólunk a munka további tervezett lépéseiről is.

2 A korpusz és a szótárak létrehozása

A munka első szakaszában létrehoztuk a vizsgálathoz szükséges szépirodalmi korpuszt, valamint az orosz nyelvű emóció- és szentimentszótárakat.

Az emóciószótárunk nyolc emóciókategóriát foglal magában, míg a szentimentlexikon egy, a negatív és egy, a pozitív értékelő tartalommal rendelkező szentimentkifejezéseket tartalmazó listát. Az emóciókifejezéseket egy orosz nyelvű monográfia mellékletére támaszkodva gyűjtöttük össze [3]. A szentimentszótárak esetében a SentiRus [4] orosz nyelvű, valószínűségi mutatókat tartalmazó lexikon listáját dolgoztuk fel, rendszereztük, valamint egészítettük ki az emóciószótárunk anyagával [1].

A korpusz három alkorpuszra oszlik, a szövegek keletkezési ideje alapján. A korpusz szövegeit automatikus módszerekkel gyűjtöttük le egy szabadon hozzáférhető műveket tartalmazó honlapról¹.

2.1 A vizsgálati korpusz létrehozása

A szövegkorpusz létrehozásának első lépése a források feltérképezése, majd a legmegfelelőbb forrás² kiválasztása volt. Választásunk egyik fontos szempontja az volt, hogy az oldal lehetővé tegye a szövegek txt formátumú automatikus gyűjtését.

A szövegek kiválasztásánál a korpusz előre meghatározott paramétereit követtük. Fő kiindulási pontnak az 1917-es orosz forradalmat választottuk. Hipotézisünk szerint ugyanis az eseményre a szépirodalomban is reflektáltak, és ezek a hatások a szövegek szintjén is kimutathatóak lesznek. Figyelembe véve a forradalom, valamint az annak hatására született művek közötti időrést, a szövegek határának az 1895-1955 közé eső időszakot választottuk. Ezt a későbbiekben további három periódusra osztottuk: az 1. korszakot 1895-1910, a 2. korszakot 1910-1920, a 3. korszakot pedig 1920-1955 közé helyeztük. Az 1. periódus a forradalom előtti, a 3. periódus a forradalom utáni

¹ <http://www.royallib.com>

² <http://www.royallib.com>

időszakot reprezentálja, míg a 2. periódus a forradalom hatására vagy az azt közvetlen megelőző években létrehozott szövegeket foglalja magába.

A létrehozott adatbázis minden szöveg esetében tartalmazza annak címét, szerzőjének nevét, valamint a szöveg keletkezésének a dátumát.

A munka gyűjtési fázisában fő célunk egy minél nagyobb adatbázis létrehozása volt. A szövegek kiválasztásánál tehát a munka e szakaszában nem követtünk semmilyen kritériumot a keletkezés időpontján kívül.

A szövegeket automatikus módszerrel gyűjtöttük, txt formátumban és UTF-8 karakterkódolással, hogy azok előfeldolgozó lépések nélkül feleljenek meg az elemző eszközök bemeneti kritériumainak. A mondatra és a szavakra bontáshoz az NLTK csomag alapeszközeit [5], a POS taggeléshez és a morfológiai elemzéshez pedig egy kimondottan orosz nyelvű szövegekre fejlesztett programot, a Pymorphy2-t használtuk [6]. A feldolgozás eredményeként megkaptuk a szövegek szavainak kisbetűs tokenjét, lemmáját és a POS-taget. Az output fájl tartalmazott néhány szisztematikus hibát (pl. az idézőjelben szereplő vagy a latin szavak felismerésének a hiánya következtében), de azokat manuális korrigáltuk.

A létrehozott korpusz adatait az 1. táblázat mutatja be.

	1. Periódus (1895-1910)	2. Periódus (1910-1920)	3. Periódus (1920-1955)	Összesen
Szöveg	46	48	82	176
Szerző	10	15	22	51
Átlagos szövegenkénti szószám	35 516	18 666	34 956	30 660
Összes szószám	1 633 741	895 978	2 866 470	5 396 189

táblázat.1. A korpusz adatai.

Ahogy a táblázatban látjuk, az adatbázisban összesen 51 szerző 176 szövege található, melyek közül legtöbb, összesen 22 szerző 82 szövege a 3. periódusban keletkezett, hiszen az öleli fel a leghosszabb időintervallumot. Ugyanakkor érdemes felhívni a figyelmet arra, hogy a 2. periódus átlagos szószáma jóval alacsonyabb, mint az egész korpuszra vonatkozó átlagos érték, míg az 1. és a 3. periódusban ugyanez valamivel magasabb. Ebből arra lehet következtetni, hogy az 1917-es forradalmat közvetlen megelőző és követő időszakban összességében rövidebb művek születtek.

2.2 Az emóció- és a szentimentszótár létrehozása

Az emóciószótárak létrehozásában egy orosz nyelvű monográfia mellékletére támaszkodtunk [3]. A melléklet összesen 8500 lexémát tartalmaz, amelyek a szerző alapján az orosz nyelvben az emóciók kifejezésére képes elemek.

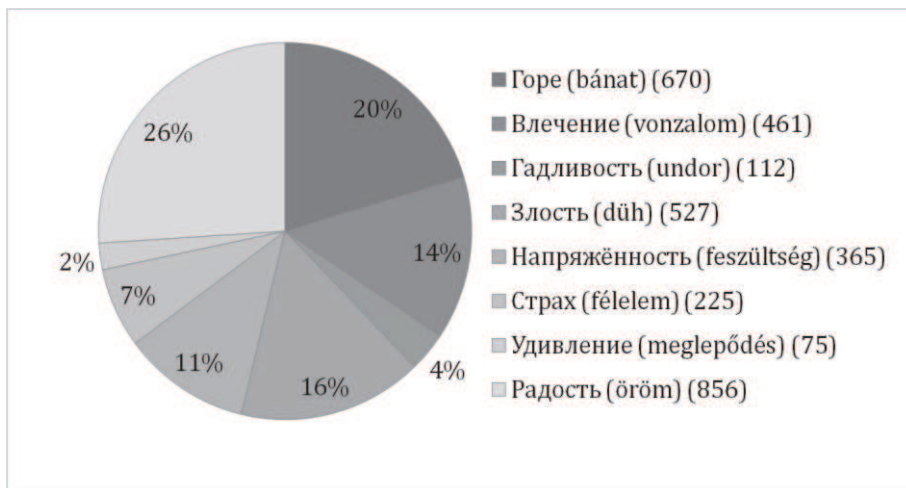
A monográfia [3] 37 emotív jelentést, azon belül számos további alkategóriát, valamint sajátos szemantikai tartalmi típusokat különít el, és ebben a komplex rendszerben dolgozza fel a lexikai anyagot. A kiinduló emotív jelentéseken belül

felsorolt lexikai elemeket 7 további kategóriába sorolja (érzelmi állapot, érzelmi állapot kialakulása, érzelmi hatás, érzelmi viszonyulás, érzelmek külső kifejeződése, érzelmi jellemzés és érzelmi minőség). A jelentések típusait a következő alkategóriák szerint különbözteti el: fő,- alapjelentés, nominatív és metaforikus, valamint képzett jelentés.

Saját szótáraink kialakításánál az első és talán legfontosabb dilemmának az érzelemkategóriák meghatározása bizonyult, hiszen erről a kérdéssel nincs egységes vélekedés sem a hazai, sem a külföldi irodalomban [7,8]. Tekintettel arra, hogy a kutatási tervek között szerepelt az angol és magyar nyelvű szövegekkel történő kontrasztív vizsgálat, az ezeken a nyelveken már meglévő és számunkra elérhető szótárak [8,9,10,11,12] felépítésére támaszkodtunk, és egy nyolc érzelemkategóriából álló séma mellett tettük le a voksunkat. A nyolc érzelemkategória az öröm, a düh, a bánat, a félelem, a feszültség, a vonzalom, az undor és a meglepődés volt [9].

A kiinduló anyagként szolgáló lista [2] feldolgozását automatikus és manuális módon végeztük el [1]. A mellékletet digitalizáltuk, majd az anyagot manuálisan javítottuk és az elemeket listákba rendeztük. A szótár kialakításakor a szófaj kritériumát tágra értelmeztük, így minden, érzelemt kifejező, vagy annak meglétére utaló elemet felvettünk a szótárba.

A kialakított szótár adatait és a kategóriák közötti megoszlási arányait az 1. ábra mutatja be.



ábra. 1. Az érzelemszótárak adatai és megoszlási arányai.

A szentimentszótár létrehozásához első lépésként kiválasztottunk egy 5000 elemet tartalmazó, orosz nyelvű lexikont [4]. A lexikont a szerzők a következőképpen készítették el: A munka első fázisában gyűjtöttek egy kisméretű, pozitív és negatív szentimentértékű elemekből álló listát, amelyeket egy korpuszra illesztettek, majd lekérdezték a szótár elemeinek környezeteit. Az eredmények alapján a korpusz minden szavához, amely szerepelt valamelyik környezetben, egy nullától egyig terjedő valószínűségi értéket rendeltek. A szám azt mutatja, hogy mekkora valószínűséggel

hordoz pozitív vagy negatív értéket az adott elem. A létrehozott szótár tehát azon a feltevésen alapul, hogy minél gyakrabban szerepel egy adott elem pozitív vagy negatív környezetben, annál nagyobb valószínűséggel lesz maga is azonos értékű (a módszer problémáiról részletesebben l. [4]).

Részben e valószínűségi mutatókra támaszkodva határoztuk meg, hogy a listából mely elemek kerüljenek be a szótárainkba. A 0.250-nél alacsonyabb valószínűségi mutatóval rendelkező elemeket eleve nem vettük figyelembe, a többi elemet pedig egyenként megvizsgáltuk, és, amennyiben elfogadtuk azt, besoroltuk a pozitív vagy a negatív lista valamelyikébe. Ezután a szentimentszótárakat az emóciószótár anyagával tovább bővítettük, ugyancsak egyenként vizsgálva a lehetséges szentimentkifejezéseket. E bővítési megoldás indokolható, hiszen, bár az emócióelemzés az érzelmek, a szentimentelemzés pedig a nyelvi értékelések azonosítását célozza, a két jelenség szorosan összefügg egymással, számos emóciókifejezés egyben értékelést is hordoz [9]. Az elkészített szentimentlexikon adatait a következő táblázat mutatja be.

Szentimentkifejezés	Elemszám
Pozitív	1982
Negatív	2675
Összes	4657

táblázat.2. A szentimentlexikon adatai.

3 Az emóció- és a szentimentelemzés eredményei

Ahhoz, hogy az emóció- és a szentimentelemzéssel olyan görbéket kaphassunk, amelyek időrendben mutatják a szövegek cselekményét, minden szöveget 100 részre osztottunk, majd a szótáras elemzéssel minden részre kiszámoltuk az emóció- és a szentimentértékeket. Az eredményeket egy interaktív dashboardon vizualizáltuk, ami lehetővé teszi az adathalmazon belüli keresést korszakonként, művenként, részenként, és megmutatja a különböző emóció- és szentimentértékeket a kiválasztott fentebbi feltételek szerint. Az alábbi ábra a negatív szentimentek előfordulásának vizualizációját mutatja be az interaktív dashboardon.



ábra. 2. Az interaktív dashboard egy részlete.

A bal oldalon lévő szentimentgörbék a korpuszon belüli szentimentértékeket mutatják. A 3 egymás alatt vízszintes görbe a 3 alkorpuszt reprezentálja, felülről alulra a következő sorrendben: 1. Periódus, 2. Periódus, 3. Periódus. Balról jobbra haladva látjuk a korábban 100 részre osztott szövegek szentimentértékeit, követve azok cselekményét. A dashboard ezen részlete így az összes szöveg szentimentértékét tartalmazza, jelen ábrán nem szűkítettük a keresést konkrét művekre. A 2. ábra jobb oldalán a harminc leggyakoribb szentimentkifejezés látható, valamint azok korszakokonkénti előfordulása. Emellett feltüntettük azt is, hogy korszakokonként összesen hány negatív értékítéletet kifejező szentimentkifejezést sikerült beazonosítani (1. Periódus: 1622, 2. Periódus: 1490, 3. Periódus: 1818).

Az emóció- és szentimentértékeket, az egész korpuszban és korszakokonkénti megoszlásban is, a 3. táblázatban tüntettük fel. Mivel a korszakokat reprezentáló alkorpuszok szószáma eltér, az emóció- és szentimentkifejezések előfordulási számát a teljes korpusz és az alkorpuszok szószámához arányítva is bemutatjuk, valamint feltüntetjük a korpuszhoz viszonyított százalékos megoszlást is. A szentimentkifejezéseknél a százaléktételeket a pozitív és a negatív szentimentkifejezésekre együtt értjük.

	1. Periódus (1895-1910)	2. Periódus (1910-1920)	3. Periódus (1920-1955)	Összesen
Összes szószám	1 633 741	895 978	2 866 470	5 396 189
Szentimentkifejezések száma	2894: 0,18%	2612: 0,29%	3228: 0,11%	8834: 0,16%
Ebből pozitív	1272	1222	1410	3904
Ebből negatív	1622	1490	1818	4930
Emóciókifejezések száma	2091: 0,13%	1843: 0,21%	2350: 0,09%	6284: 0,12%

táblázat.3. Emóció- és szentimentértékek a NarRu korpuszban.

Ahogy az előfordulási arányok mutatják, az emóció- és a szentimentkifejezések legnagyobb százalékban a 2. periódusban fordulnak elő, annak ellenére, hogy ez öleli fel a legrövidebb időintervallumot. Míg az 1. periódus hasonló adatai alapvetően egyeznek az egész korpuszra számított értékekkel, a legnagyobb eltérés a 2. és a 3. periódus között mutatkozik. Utóbbiban ugyanis csupán kis százalékban voltak beazonosíthatók az emóció- és a szentimentkifejezések.

Az emócióelemzés során azt az eredményt kaptuk, hogy az emóciókifejezések emóciókategóriánkénti megoszlása összhangban áll az emóciószótár elemeinek emóciókategóriánkénti százalékos megoszlásával, ezért azt külön táblázatban nem reprezentáljuk.

4 A szövegek narratívaelemzése

A narratívaelemzés alapját a 100 részre feldarabolt szövegek egyes részenkénti szentimentértéke adta.

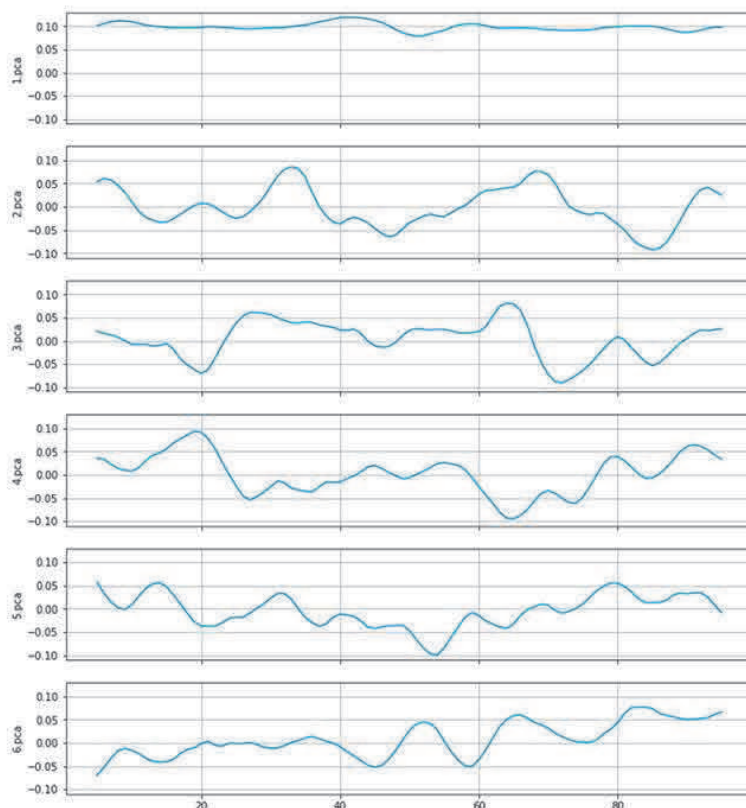
A kapott értékekből főkomponens elemzéssel nyertük ki a korpuszban domináló narratívagörbét. Az elemzést elvégeztük a szentimentértékek összességére és átlagára is. Az irodalomban talán jellemzőbb a szentimentértékek összegének alkalmazása [2], miután az átlagok számítása révén veszítünk adataink változatosságából.

A szentimentértékek összességét nézve 6, valamint 12 főkomponenst nyertünk ki az adatokból. Előbbi esetben az adataink kezdeti varianciájának kicsivel több mint 50, utóbbi esetben majdnem 60%-át át őriztük meg. A szentimentértékek átlagából 6 főkomponenst különítettünk el, és adataink kezdeti varianciájának 80%-át őriztük meg.

A kapott narratívagörbéket mozgóátlagot számolva kisimítottuk. Ehhez az adott időpont előtt és utána lévő 5 szövegrész szentimentértékének átlagát vettük figyelembe, kisebb súllyal számolva azokat az értékeket, melyek távolabb esnek az adott szövegrésztől.

Pearson-féle korrelációt számoltunk, hogy kiderüljön, mely művek simított narratívagörbéi állnak legközelebb az egyes narratívátípusokhoz.

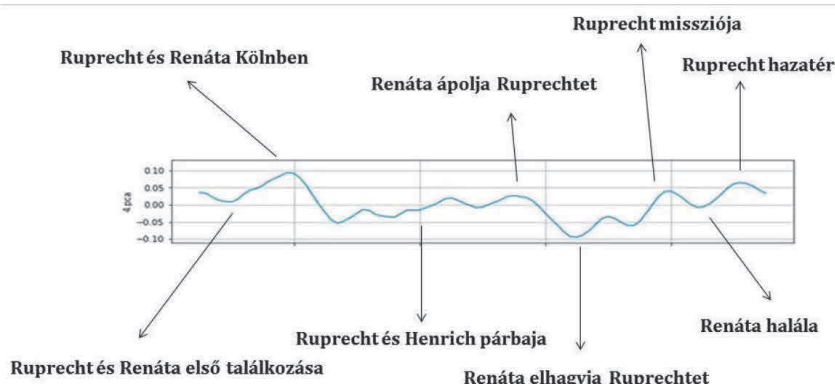
A kapott narratívátípusokat összevetettük Reagan és szerzőtársai [2] angol nyelvű szövegek elemzésével kapott narratívagörbéivel, amelyek a szentimentértékek összességére számolt narratívákkal mutatták a nagyobb egyezést. A szentimentértékek összességére végzett elemzésből az alábbiakban az áttekinthetőség kedvéért a 6 főkomponenst megkülönböztető eredményeket prezentáljuk:



ábra. 3. A szentimentek összesített értékeiből definiált 6 főkomponens.

A 6 főkomponens közül az 1. egy pozitívan stagnáló görbe, a legegyszerűbb cselekményszerkezetnek tekinthető. Az 5. görbe egy 'u' alakú verzió, ahol a történet közepén lévő mélypont után visszatérünk a kezdeti állapothoz, míg a 6. főkomponens egy emelkedő narratívagörbe. A 2. és a 3. görbe mutatja a legbonyolultabb cselekményt. A 4. főkomponens az 5.-hez hasonlóan visszatér a kiinduló ponthoz, annyi különbséggel, hogy a történetben két, egy szolidabb és egy drasztikusabb mélypont mutatható ki, amelyek között egy enyhe pozitív felemelkedés látható.

Narratívagörbéiket Reagan és szerzőtársai [2] minden esetben két verzióra osztották az alapján, hogy az pozitív vagy negatív véggel záródik. A kutatás során ezt a felosztást nem alkalmaztuk, de amennyiben ebből a szempontból tekintjük saját eredményeinket, azt látjuk, hogy azok, az 5. főkomponenstől eltekintve, mindegyik főkomponens inkább pozitív értékkel záródik, bár a 2. és a 4. görbe a történet végén enyhe esést mutat. A következő ábra Valerij Brjusov Tüzes angyal c. művének cselekményét mutatja be. A történet a 4. főkomponenshez kapcsolható 0.783123 korrelációs értékkel. A narratívagörbén a történet fordulópontjait és főbb eseményeit jelöltük be.



ábra. 3. V. Brjusov Tüzes angyal c. műve.

A regény cselekménye a 4. főkomponenshez tartozik. A történet vége visszatér a kiinduló ponthoz, pozitív értékkel kezdődik és azzal is záródik. A pozitív kezdés a két kulcsfigura találkozása és kettejük Kölnben töltött ideje köré szerveződik, míg a pozitív vég a főszereplő hazatérésére és missziójára utal. A 4. főkomponens két mélypontja is magyarázható. A szolidabb mélypont egy párbajjal és annak negatív kimenetével, a drasztikusabb pedig a két szereplő elválásával. A mélypontok közötti enyhe pozitív felemelkedés a sérült párbaj utáni ápolásával köthető össze.

5 Összegzés

A tanulmányban egy olyan kutatásról számoltunk be, amelynek célja az volt, hogy különböző orosz nyelvű szépirodalmi szövegek narratívaserkezetét szentiment- és emóciószótárak segítségével feltárjuk, valamint a legfőbb narratívaserkezeteket definiáljuk. A kutatáshoz minden vizsgálati és elemzési eszközt magunk hoztunk létre [1], amelyek fontos adalékul szolgálhatnak a jövőbeli, orosz nyelvre irányuló tartalomelemzési feladatainkhoz is.

A kutatás első lépéseként létrehoztunk egy összesen 51 író 176 szövegét tartalmazó szépirodalmi korpuszt. Annak céljából, hogy a különböző korszakokban keletkezett művek struktúráját összevethessük egymással, az adatbázisban három alkorpuszt különítettünk el.

Az első feldolgozási szakaszban a korpuszt mondatokra és szavakra bontottuk, amihez az NLTK csomag alapeszközeit használtuk [5]. A szövegek POS taggeléséhez és a morfológiai elemzéséhez egy orosz nyelvű szövegekre fejlesztett programot, a Pymorphy2-t alkalmaztuk [6]. Az ily módon feldolgozott korpuszt ezután az emóció- és szentimentszótárunkkal, szótárillesztéses megoldással elemeztük. A szótáras elemzésekkel megkaptuk a korpuszban bemutatott korszakokra jellemző szisztematikus mintázatokat az emóciók és a szentimentek vonatkozásában.

A szótárakat ugyancsak e kutatás keretében, magunk hoztuk létre. Az emóciósztárunk, amelyben nyolc érzelmetípust különböztetünk meg, összesen 3291 elemet tartalmaz. A szentimentsztár egy negatív és egy pozitív értékelő nyelvi elemeket tartalmazó listából áll és összesen 4657 szentimentkifejezést tartalmaz.

A kapott eredmények alapján feltártuk a szövegek narratívastruktúráit, valamint megállapítottuk, mely szövegek tartoznak a különböző szerkezet típusokhoz. Végül az eredményeket egy interaktív dashboard segítségével vizualizáltuk. A dashboard nyújtotta lehetőségeket szeretnénk még jobban kiaknázni, és azt további kutatásokra és elemzésekre is felhasználni.

Az elkészített korpusz, a vizsgálati eszközök, valamint az interaktív dashboard a Precognox Informatikai kft. TAS elnevezésű nyelvfeldolgozó rendszerének részét képezi (<http://precognox.hu/tas-szoveganalitikai-rendszer>). A dashboardot hamarosan kutatási célra szabadon hozzáférhetővé tesszük a honlapon.

Köszönetnyilvánítás

A kutatást a Precognox Informatikai Kft., az Emberi Erőforrások Minisztériuma Új Nemzeti Kiválóság Programja, valamint az Európai Kutatási Tanács (European Research Council) az Európai Unió Horizont 2020 kutatási és innovációs programjának keretében (ERC_CoG_2014_648693 sz. szerződésben) támogatta.

Bibliográfia

1. Nyíri, Zs.: Egy orosz nyelvű emóció- és szentimentsztár, valamint egy orosz nyelvű szépirodalmi korpusz létrehozásának tapasztalatai. In: II. Szláv Filológiai Konferencia, Budapest (2017) Megjelenés előtt
2. Reagan, A. J., Mitchell, L., Kiley, D., McLaughlin Danforth, C., Dodds, P. S.: The emotional arcs of stories are dominated by six basic shapes. In: EPJ Data Science (2016) Elérhető: <https://arxiv.org/pdf/1606.07772v2.pdf>
3. Бабенко, Л. Г.: Лексические средства обозначения эмоций в русском языке. Издательство Уральского университета, Екатеринбург (1989)
4. Chetviorkin, I. I., Loukachevitch, N. V.: Extraction of Russian Sentiment Lexicon for Product Meta-Domain. Proc. of COLING 2012: Technical Papers (2012) 593–610
5. Bird, S., Klein, E., Loper, E.: Natural Language Processing with Python, Analyzing Text with the Natural Language Toolkit, O'Reilly Media (2009) Elérhető: http://www.nltk.org/book_1ed/
6. Морфологический анализатор Pymorphy2 (2015) Elérhető: <https://pymorphy2.readthedocs.io/en/latest/>
7. Liu, B.: Sentiment Analysis and Opinion Mining. Draft (2012) Elérhető: <https://www.cs.uic.edu/~liub/FBS/SentimentAnalysis-and-OpinionMining.pdf>
8. Szabó, M. K., Morvay, G.: Emócióelemzés magyar nyelvű szövegeken. In: Gecső, T., Sárdi, Cs. (eds.): Nyelv, Kultúra, Társadalom, Tinta Könyvkiadó, Budapest (2015) 286-293
9. Szabó, M. K., Vincze, V., Morvay, G.: Magyar nyelvű szövegek emócióelemzésének elméleti és nyelvtéchnológiai problémái. In: Reményi, A. Á., Sárdi, Cs., Tóth, Zs. (eds.):

- Távlatok a mai magyar alkalmazott nyelvészetben, Tinta Könyvkiadó, Budapest (2016) 282-293
10. Szabó, M. K.: Egy magyar nyelvű szentimentlexikon létrehozásának tapasztalatai és dilemmái. In: Gecső T., Sárdi, Cs. (eds): *Nyelv, Kultúra, Társadalom*, Tinta Könyvkiadó, Budapest (2015) 278-286
 11. Wilson, T., Wiebe J., Hoffmann P.: Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis. Proc. of HLT-EMNLP-2005 (2005) Elérhető:
 12. Strappava, C., Mihalcea. R.: Learning to identify emotions in text. Elhangzott: UMM CSci Senior Seminar Conference, University of Minnesota, Morris. Morris, Amerikai Egyesült Államot (2008) Elérhető:
<http://citeseerx.ist.psu.edu/viewdoc/download?rep=rep1&type=pdf&doi=10.1.1.217.62>

Forrás

<http://www.royallib.com>

<http://precognox.hu/tas-szoveganalitikai-rendszer>